WHAT IS CLAIMED IS:

1. A load balancer that collects server capability information for a plurality of servers, wherein the server capability information is based at least in part on processing of sample requests transmitted to the plurality of servers during intervals, and that load balances client requests in accordance with the collected server capability information.

2. The load balancer of claim 1 that encodes the collected server capability information to represent the plurality of servers in accordance with proportional server capability of each of the plurality of servers.

3. The load balancer of claim 2 wherein the load balancing of client requests comprises selection of entries from the proportional server capability encoding, wherein each entry indicates at least one of the plurality of servers.

4. The load balancer of claim 3 wherein the selection of entries is random or pseudo-random.

5. The load balancer of claim 3 wherein the selection of entries is predetermined.

6. The load balancer of claim 5 wherein the selection of entries is sequential.

7. The load balancer of claim 1 wherein the collecting of server capability information comprises transmitting the sample requests to the plurality of servers during the intervals and recording information that corresponds to the servers servicing of the sample requests.

8. The load balancer of claim 1 wherein the sample requests include a mixture of configurable directory requests.

9. The load balancer of claim 1 wherein the server capability information includes one or more of proportion of serviced sample requests, time to serve each

sample request, time to serve total sample requests, proportion of sample request types serviced, and time to serve sample request types.

10. The load balancer of claim 1 that updates a proportional server capability based load balancing encoding in accordance with the collected server capability information.

11. The load balancer of claim 10 that updates the proportional server capability based load balancing encoding in response to a change in network configuration.

12. The load balancer of claim 11 wherein the change of network configurations includes change of server availability.

13. The load balancer of claim 1 embodied in one or more of cache, registers, memory, and fast look-up tables.

14. A method comprising:

during intervals, collecting data that reflects capabilities of a plurality of backend servers, wherein the backend server capability data is based at least in part on servicing of sample requests by the plurality of backend servers; and

encoding the collected backend server capability data to reflect proportional backend server capability of each of the plurality of backend servers.

15. The method of claim 14 wherein the collected backend server capability data is encoded to indicate each of the plurality of backend servers in accordance with their proportional capability based at least in part on the collected backend server capability data.

16. The method of claim 14 further comprising updating the encoding in accordance with the collected data.

17. The method of claim 14 further comprising load balancing client requests in accordance with the encoding.

18. The method of claim 17 wherein the load balancing comprises randomly selecting entries from the encoding, wherein the encoding includes entries that indicate the plurality of backend servers.

19. The method of claim 17 wherein load balancing comprises selecting predetermined entries from the encoding, wherein the encoding includes entries that indicate the plurality of backend servers.

20. The method of claim 19 wherein the predetermined selection of entries is sequential.

21. The method of claim 14 wherein the collected backend server capability data includes one or more of proportion of sample requests serviced by each of the backend servers, time for each of the backend servers to serve each sample request, time for each of the backend servers to serve total sample requests, proportion of sample request types serviced by each of the backend servers, and time for each of the backend servers to serve sample request types.

22. The method of claim 14 wherein collecting backend server capability data comprises:

transmitting the sample requests to the backend servers; and

recording data that corresponds to servicing of the sample requests by the backend servers.

23. The method of claim 14 wherein the sample requests include sample directory requests.

24. The method of claim 23 wherein the sample directory requests are in accordance with one or more protocol including lightweight data access protocol and universal description discovery and integration.

25. The method of claim 14 embodied as a computer program product encoded on one or more machine-readable media.

26. A method comprising:

load balancing client requests across a plurality of servers in accordance with a proportional server capability information encoding that reflects proportional capabilities of the plurality of servers,

wherein the reflected proportional server capability information is based at least in part on servicing of sample requests by the plurality of servers.

27. The method of claim 26 wherein encoding reflects frequency of sample requests serviced by the servers.

28. The method of claim 27 wherein the frequency of sample requests serviced includes one or more of number of sample requests serviced during a time interval, number of sample requests serviced during a time interval based on type of sample requests, time to service a number of sample requests, and time to service a number of sample requests based on type of sample requests.

29. The method of claim 26 further comprising maintaining the proportional server capability information encoding.

30. The method of claim 26 further comprising the proportional server capability information at intervals between servicing of client requests.

31. The method of claim 29 wherein maintaining the proportional server capability information encoding comprises:

transmitting the sample requests to the plurality of servers at intervals; and

recording the server capability information that indicates frequency of the sample requests serviced by the servers.

32. The method of claim 31 wherein the sample requests include a mixture of configurable sample requests.

33. The method of claim 26 wherein the encoding includes a data structure that indicates the plurality of servers in accordance with the proportional server capability information.

34. The method of claim 33 wherein the load balancing comprises selecting entries from the data structure at random.

35. The method of claim 34 wherein the load balancing comprises predetermined selection of entries from the data structure.

36. The method of claim 26 embodied as a computer program product encoded in one or more machine-readable medium.

37. A method comprising:

during a data collection interval,

transmitting sample requests to servers,

recording data that corresponds to servicing of the transmitted sample requests by each of the servers; and

encoding the recorded data, wherein the encoding of the data indicates each of the servers in accordance with their proportional server capability based at least in part on the recorded data.

38. The method of claim 37 wherein the encoding includes a load balancing table.

39. The method of claim 38 further comprising randomly selecting entries from the load balancing table to load balance client requests across the servers.

40. The method of claim 38 further comprising predetermined selection of entries from the load balancing table to load balance client requests across the servers.

41. The method of claim 37 further comprising load balancing client requests in accordance with the encoding.

42. The method of claim 37 further comprising randomly selecting entries from the load balancing structure to load balance client requests.

43. The method of claim 37 wherein the sample requests include search requests, compare requests, and update requests.

44. The method of claim 37 wherein the recorded data indicates one or more of number of sample requests serviced during the data collection interval by each of the directory servers, number of sample requests serviced during the data collection interval by each of the directory servers based on sample request type, time for each directory server to service a number of sample requests during the data collection interval, and time for each of the directory servers to service a number of sample requests based on type of sample requests during the data collection interval.

45. The method of claim 37 further comprising servicing client requests with a second plurality of servers during the data collection interval.

46. The method of claim 37 further comprising buffering client requests during the data collection interval.

47. The method of claim 37 embodied as a computer program product encoded in one or more machine-readable medium.

48. An apparatus comprising:

a network interface; and

means for load balancing client requests in accordance with a proportional server capability load balancing information encoding that is updated in accordance with server capability information based at least in part on processing of sample requests by a plurality of servers during intervals.

49. The apparatus of claim 48 further comprising a high resolution timer for measuring the server capability information.

50. The apparatus of claim 48 further comprising means for collecting the server capability information at intervals between handling of client requests.

51. The apparatus of claim 48 further comprising means for measuring the server capability information.

52. The apparatus of claim 48 further comprising memory that includes the proportional server capability load balancing information encoding.

53. A computer program product encoded in one or more machine-readable media, the computer program product comprising:

a first sequence of instructions to transmit sample requests to a plurality of servers at intervals and receive responses corresponding thereto; and

a second sequence of instructions to determine proportional capability information for each of the plurality of servers based at least in part on the sample requests and corresponding responses.

54. The computer program product of claim 53 further comprising a third sequence of instructions to encode the determined proportional capability of each of the plurality of servers, wherein the encoded proportional server capability indicates each of the plurality of servers in accordance with the determined proportional server capability.

55. The computer program product of claim 54 further comprising the third sequence of instructions to maintain the proportional server capability encoding for load balancing.

56. The computer program product of claim 54 further comprising a fourth sequence of instructions to load balance client requests in accordance with the proportional server capability encoding.

57. The computer program product of claim 56 further comprising a fifth sequence of instructions to buffer client requests during the intervals.

58. The computer program product of claim 56 further comprising the third sequence of instructions to forward client requests to standby servers during the intervals.

59. The computer program product of claim 53 wherein the second sequence of instructions measures one or more of time for each of a plurality of servers to respond to each request, time for each of a plurality of servers to respond to each request based on request type, total number of responses provided by each of a plurality of servers during the periodic intervals, and number of responses provided by each of a plurality of servers based on request type.

60. A computer program product encoded in one or more machine-readable media, the computer program product comprising:

a first sequence of instructions to update a proportional server capability load balancing information encoding that reflects proportional measured sample request based capabilities of a plurality of servers, wherein the capabilities are measured during intervals; and

a second sequence of instruction to select server indications from the proportional server capability load balancing information encoding to load balance client requests.

61. The computer program product of claim 60 further comprising a third sequence of instructions to buffer client requests while the first sequence of instructions updates the proportional server capability load balancing information encoding.

62. The computer program product of claim 60 further comprising a third sequence of instructions to forward client requests to standby servers while the first sequence of instructions updates the proportional server capability load balancing information encoding.

63. The computer program product of claim 60 further comprising a third sequence of instructions to measure capabilities of the plurality of directory servers based at least in part on sample requests.

64. The computer program product of claim 63 wherein the third sequence of instructions to measure capabilities comprises the third sequence of instructions to transmit the sample requests to the plurality of directory servers and to receive responses corresponding thereto during the intervals.

65. The computer program product of claim 60 wherein the measured sample request based proportional capabilities include one or more of number of sample directory requests serviced during a time interval, number of sample directory requests serviced during a time interval based on type of sample requests, time to service a number of sample directory requests during a time interval, and time to service a number of sample directory requests based on type of sample requests during a time interval.

66. The computer program product of claim 60 wherein the requests include directory requests.

67. A network comprising:

a plurality of servers processing requests; and

a load balancer forwarding client requests in accordance with a proportional server capability information encoding that indicates each of the plurality of servers in accordance with their proportional capability, wherein the proportional server capability information encoding is based at least in part on servicing of sample requests during intervals between forwarding of client requests.

68. The network of claim 67 further comprising the load balancer measuring the capabilities of the plurality of servers during the intervals.

69. The network of claim 68 wherein measuring includes the load balancer transmitting the sample requests and receiving responses corresponding to the sample requests.

70. The network of claim 67 further comprising one or more standby servers to handle client requests while the load balancer measures capabilities of the plurality of servers.